# Distributed Hypothesis-Testing on Graphs: Algorithms and Convergence Rates

**Angelia Nedić**

Collaborative work with **Alexander Olshevsky and Cesar Uribe**

School of Electrical Computer and Energy Engineering
Arizona State University at Tempe

**April 13, 2017**

# Centralized Case

# Bayes' Rule Belongs to Stochastic Approximations

# Bayesian Learning - Hypothesis Testing

- An agent observes a certain phenomenon, and uses the observations to refine its understanding of the state of the phenomenon.

- The observation sequence $\{S_k\}$ is random i.i.d., drawn according to an (unknown) distribution $f$, with $S_k \in \mathcal{S}$ for all $k$

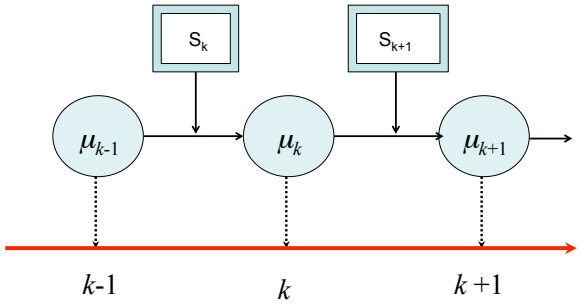
- Let $\mathcal{S}$ be a finite set of the possible states for the phenomenon.

- An agent, receives realizations $s_1, s_2, \ldots$ of the true state in discrete time instances $k = 1, 2, \ldots$

- Agent has a set of hypothesis $\{\ell_1(\cdot), \ldots, \ell_m(\cdot)\}$ of probability distributions on $\mathcal{S}$, and would like to determine a hypothesis that the $best$ describes the data collected over time, i.e., determine an $\ell_i$ that is the closest to $f$ in some sense

# Bayes' Update: Inference Via Minimization Rule

- Initial distribution $\mu_0$ is selected at time $t = 0$.
- At time $k$, agent has a belief $\mu_k$ (probability distribution on $\{1, 2, \ldots, m\}$) that the best explains the observations $s_1, \ldots, s_k$ collected up to that time. At time $k+1$, it observes $s_{k+1}$ and updates its belief to $\mu_{k+1}$:




Bayesian update: for all $i = 1, \ldots, m$,

$$\mu_{k+1}^i = \frac{\mu_k^i \ell_i(s_{k+1})}{\sum_{p=1}^m \mu_k^p \ell_p(s_{k+1})}$$

- Bayes' rule minimizes a function composed of two terms (Walker 2006, Zellner 1988)
  - Maximum Likelihood Estimation (MLE) of a state given the observed data and
  - A regularization function Kullback-Leibler (KL) divergence from the current prior

$$\mu_{k+1} = \underset{\pi \in \Delta_m}{\operatorname{argmin}} \left\{ \underbrace{-\langle \ln \ell(s_{k+1}), \pi \rangle}_{MLE} + D_{KL}(\pi \| \mu_k) \right\}$$

$\Delta_m$ is the set of $m$-dim. probability distributions, $\ell(s_{k+1}) = [\ell_1(s_{k+1}), \ldots, \ell_m(s_{k+1})]'$, and $D_{KL}(p\|q) = \sum_{i=1}^m p_i \ln \frac{p_i}{q_i}$.

# Digression: Gradient-Projection/Mirror-Descent Method

- Minimize convex function $F : \mathbb{R}^m \to \mathbb{R}$ subject to closed convex set $X \subset \mathbb{R}^m$

- $X$ is assumed simple (the projection on the set is easy)

- Euclidean-norm, gradient projection

$$x_{k+1} = \Pi_X \left[x_k - \alpha_k \nabla F(x_k)\right] \iff x_{k+1} = \underset{y \in X}{\operatorname{argmin}} \left\{ \langle \nabla F(x_k), y \rangle + \frac{1}{2\alpha_k} \|y - x_k\|^2 \right\}$$

$\Pi_X(y)$ is the Euclidean projection of $y$ on $X$ and $\alpha_k > 0$ is a stepsize.

- Extension from the use of the Euclidean norm to the use of a Bregman function: **Mirror-descent method**

$$x_{k+1} = \underset{y \in X}{\operatorname{argmin}} \left\{ \langle \nabla F(x_k), y \rangle + \frac{1}{\alpha_k} B(y, x_k) \right\}$$

where $B(y, x)$ is a Bregman distance function

# Mirror-Descent Method using KL-divergence

**Fact** KL-divergence is a Bregman distance function on $\Delta_m$, induced by $t \mapsto t \ln t$

Hence, when $X = \Delta_m$

**Mirror-Descent Method**

$$x_{k+1} = \underset{y \in \Delta_m}{\operatorname{argmin}} \left\{ \langle \nabla F(x_k), y \rangle + \frac{1}{\alpha_k} D_{KL}(y, x_k) \right\}$$

where $B(y, x)$ is a Bregman distance function

**Bayes' Rule**

$$\mu_{k+1} = \underset{\pi \in \Delta_m}{\operatorname{argmin}} \left\{ -\langle \underbrace{\ln \ell(s_{k+1})}_{sample\ \ gradient}, \pi \rangle + D_{KL}(\pi \| \mu_k) \right\}$$

**Bayes' Rule: stochastic mirror-descent method using KL-divergence (as Bregman distance function) and a fixed stepsize!**

- What optimization problem is being solved by Bayes' Update Rule?

# Optimizing Expected Log-Likelihood

$$\mu_{k+1} = \underset{\pi \in \Delta_m}{\operatorname{argmin}} \left\{ -\langle \underbrace{\ln \ell(s_{k+1})}_{sample\ \ gradient}, \pi \rangle + D_{KL}(\pi \| \mu_k) \right\}$$

The update rule is a stochastic mirror-descent method for solving an LP

$$\text{minimize} \quad -\langle \underbrace{\mathbb{E}_f[\ln \ell(S)]}_{unknown}, \pi \rangle \quad \text{subject to} \quad \pi \in \Delta_m,$$

where $f$ is the unknown distribution of the random variable $S$, whose realization $s_{k+1}$ at time $k+1$ is used in the method.

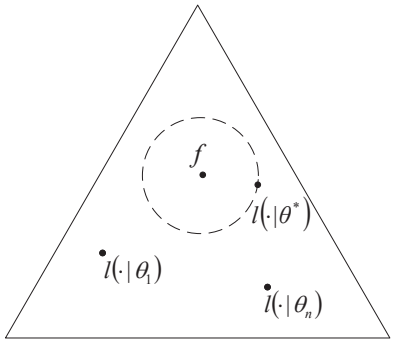
Since $f$ is unknown, we use samples of the gradients (akin to stochastic approximation methods Robbins and Monroe 1957, Polyak and Juditsky 1992)

**The problem is equivalent to another LP**

$$\text{minimize} \quad \sum_{i=1}^{m} \pi_i \underbrace{D_{KL}(f \| \ell_i)}_{unknown} \quad \text{s.t.} \quad \pi \in \Delta_m.$$

Bayes' update rule is a stochastic mirror-descent method for solving the above "uncertain" LP problem, which is equivalent to

$$\min_{1 \le i \le m} D_{KL}(f \| \ell_i)$$



# Consistency of the Bayes' update

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{m} \pi_i D_{KL}\left(f \| \ell_i\right) \\ \text{subject to} \quad & \pi \in \Delta_m. \end{aligned}$$

Equivalent to

$$\min_{1 \le i \le m} D_{KL}\left(f \| \ell_i\right)$$

Let

$$M^* = \left\{ i^* \mid D_{KL}\left(f \| \ell_{i^*}\right) = \min_{1 \le i \le m} D_{KL}\left(f \| \ell_i\right) \right\}.$$

**Assumptions**:

- $\mu_0^j > 0$ for at least one $j \in M^*$ (e.g., choose the uniform distribution initially)
- $\operatorname{supp}(f) \subseteq \operatorname{supp}(\ell_i)$ for all hypotheses $i$

**Result 1**: Under these assumptions

$$\lim_{k \to \infty} \mu_k^j = 0 \qquad \text{almost surely for all } j \notin M^*.$$

# Convergence rate of Bayes' update

**Result**: Under the preceding assumptions, we have: for any given $\rho > 0$, there is an integer $\boldsymbol{N}(\rho)$ such that with probability $1 - \rho$, we have

$$\mu_k^j \leq \exp\left\{-k\frac{\gamma_2}{2} + \gamma_1\right\} \qquad \text{for all } j \notin M^* \text{ and all } k \geq \boldsymbol{N}(\rho),$$

where

$$\boldsymbol{N}(\rho) \triangleq \left\lceil \frac{8\left(\ln\left(\alpha\right)\right)^2 \ln\left(\frac{1}{\rho}\right)}{\gamma_2^2} \right\rceil,$$

$\alpha$ is a lower bound on the likelihoods $\ell_i$ on the support of $f$:

$$\alpha = \min_{i\in[m]} \min_{j\in \mathrm{supp}(\ell_i)} \ell_i(s_j)$$

$$\gamma_1 \triangleq \max_{\substack{j\in[m]\setminus M^* \\ i^*\in M^*}} \left\{\ln \frac{\mu_0^j}{\mu_0^{i^*}}\right\} \qquad \gamma_2 \triangleq \min_{j\in[m]\setminus M^*} \left\{D_{KL}(f\|\ell_j) - D_{KL}\left(f\|\ell_{i^*}\right)\right\}$$

Note that the expression for $\gamma_1$ suggests that using the uniform prior is a good choice.

Also, if we use observations initially to form a good prior $\mu_0$, the rate is better.

$\gamma_2$ captures how well we differentiate correct from wrong - affects the rate and $\boldsymbol{N}(\rho)$.

**Rate is exponential with high probability for large enough $k$.**

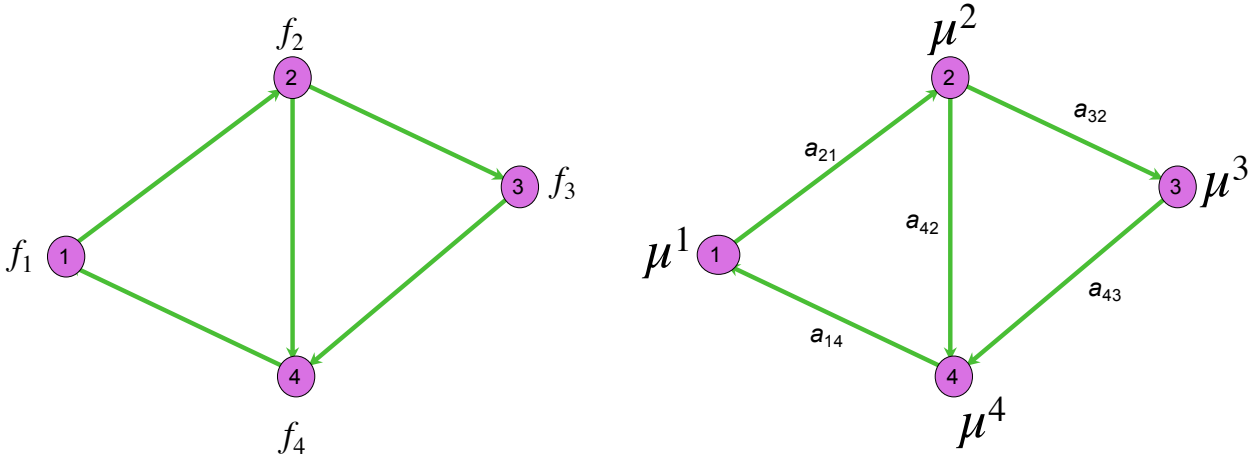
# Bayes Rule - Optimization

- Simplex method - finite convergence but for deterministic problem
- First-order stochastic methods (simple form) have a sub-linear rate as the best convergence rate
- In the case of Bayes' setting we have linear rate with high probability after some time
- What other (stochastic) optimization problems share this property?

# Decentralized Hypothesis Testing

- A system of $n$ agents, each with its own collection of distributions $\ell^i(\cdot \mid \theta)$ over a common set of $m$ hypothesis, denoted by $\Theta = \{\theta_1, \ldots, \theta_m\}$
- Each agent receives private observations $\{s_k^i\}$
- The observations of agent $i$ are realizations of an i.i.d. random sequence $\{S_k^i\}$ with $S_k^i \in \mathcal{S}^i$ drawn according to an (unknown) probability distribution $f^i$
- The set $\mathcal{S}^i$ of realizations is assumed to be finite
- Agents communicate over time varying undirected graphs $G_k, k = 1, 2 \ldots$
- Agents do not share their observations, but share their beliefs locally (with the neighbors in the graph)
- How should agents aggregate their information to ensure learning the hypothesis that explains the best the whole data they have collected?

# Decentralized Setting: How to aggregate the beliefs?




- Simply linear in beliefs: every agent updates by:
  - $\widehat{\mu}_k^i = \sum_{i=1}^n [A_k]_{ij} \mu_k^j$ (neighbors exchange beliefs)
    Here $A_k$ is a stochastic matrix compatible with the graph $G_k = ([n], E_k)$:
    $A_k$ has nonnegative entries, with $[A_k]_{ij} > 0$ when $\{i, j\} \in E_k$,
    and each row sum is equal to 1
  - Followed by a Bayes' update with the new observation $\ell^i(s_{k+1}^i|\cdot)$ to obtain $\mu_{k+1}^i$.
- It has been done by Shahrampour and Jadbabaie 2013, Jadbabaie, Molavi, and Tahbaz-Salehi 2013, 2015
- Not intuitive (axiomatic approach Jadbabaie, Molavi, and Tahbaz-Salehi 2015)
- Looking for an optimization-based approach

# Decentralized Setting: Aggregate the beliefs analogous to the single agent case

- Single agent (opt. view)
  Agent solves the problem $\min_{\theta\in\Theta} D_{KL}(f\|\ell(\cdot\mid\theta))$
  by performing Bayes' updates

$$\mu_{k+1} = \underset{\pi\in\mathbb{P}(\Theta)}{\operatorname{argmin}} \{-\langle \ln \ell(s_{k+1}\mid\cdot),\pi\rangle + D_{KL}(\pi\|\mu_k)\}$$

- Multi-agent case
  Agents want to solve $\min_{\theta\in\Theta}\sum_{i=1}^n D_{KL}(f^i\|\ell^i(\cdot\mid\theta))$ by performing

$$\mu^i_{k+1} = \underset{\pi\in\mathbb{P}(\Theta)}{\operatorname{argmin}} \left\{ -\left\langle \ln \ell^i(s^i_{k+1}|\cdot),\pi\right\rangle + \sum_{j=1}^n [A_k]_{ij} D_{KL}(\pi\|\mu^j_k) \right\}$$

  where $\mathbb{P}(\Theta)$ is the set of all probability distributions of the finite hypothesis set $\Theta$, or in a closed form: for all agents $i$,

$$\mu^i_{k+1}(\theta) = \frac{\prod_{j=1}^n \mu^j_k(\theta)^{[A_k]_{ij}}\ell^i(s^i_{k+1}|\theta)}{\sum_{r=1}^m\prod_{j=1}^n \mu^j_k(\theta_r)^{[A_k]_{ij}}\ell^i(s^i_{k+1}|\theta_r)} \qquad \text{for all } \theta\in\Theta$$

# Assumptions: Graphs, Likelihoods and Initial Beliefs

**Assumption** (Graphs) Graph sequence $\{G_k\}$ and a matrix sequence $\{A_k\}$ are such that:

1. $\{G_k\}$ is undirected and $B$-connected, i.e., there is an integer $B \geq 1$ such that the graph $\left\{V, \bigcup_{t=kB}^{(k+1)B-1} E_t\right\}$ is strongly connected for all $k \geq 0$.
2. $A_k$ is doubly-stochastic with $[A_k]_{ij} > 0$ if $\{i,j\} \in E_k$ and $[A_k]_{ii} > 0$.
3. If $[A_k]_{ij} > 0$ then $[A_k]_{ij} > \eta$ for some positive scalar $\eta$.

**Assumption** (Likelihood Models)

$$\Theta^* \triangleq \bigcap_{i=1}^{n} \operatorname*{argmin}_{\theta \in \Theta} D_{KL}\left(f^i(\cdot) \,\|\, \ell^i(\cdot|\theta)\right) \quad \text{is nonempty}$$

**Assumption** (Initial Beliefs) For all agents $i = 1, \ldots, n$,

1. The prior beliefs on all $\theta^* \in \Theta^*$ are positive, i.e. $\mu_0^i(\theta^*) > 0$ for all $\theta^* \in \Theta^*$.
2. There exists an $\alpha > 0$ such that $\ell^i\left(s^i|\theta\right) > \alpha$ for all outcomes $s^i$ and all $\theta \in \Theta$.

# Consistency

**Proposition 1** Under these assumptions, the update rule

$$\mu_{k+1}^i(\theta) = \frac{\prod_{j=1}^n \mu_k^j(\theta)^{[A_k]_{ij}\ell^i(s_{k+1}^i|\theta)}}{\sum_{r=1}^m \prod_{j=1}^n \mu_k^j(\theta_r)^{[A_k]_{ij}\ell^i(s_{k+1}^i|\theta_r)}}$$

generates sequences $\{\mu_k^i\}$, $i = 1, \ldots, n$, such that with probability 1, for all agents $i$:

$$\lim_{k\to\infty} \mu_k^i(\theta) = 0 \quad \text{for all } \theta \notin \Theta^*$$

Proof: Choose a $\theta^* \in \Theta^*$ and define

$$\varphi_k^i(\theta) = \ln\left(\frac{\mu_k^i(\theta)}{\mu_k^i(\theta^*)}\right)$$

From the update rule, we obtain for all $i = 1, \ldots, n$

$$\varphi_{k+1}^i(\theta) = \sum_{j=1}^n [A_k]_{ij}\varphi_k^j(\theta) + \ln\left(\frac{\ell^i(s_{k+1}^i|\theta)}{\ell^i(s_{k+1}^i|\theta^*)}\right) \quad \text{for all } \theta \in \Theta$$

Stacking $\varphi_{k+1}^i$, for $i = 1, \ldots, n$ in a vector

$$\varphi_{k+1}(\theta) = A_k\varphi_k(\theta) + \mathcal{L}_k(\theta) \quad \text{for all } \theta \in \Theta$$

and the rest of analysis is akin to analysis of "consensus"-based algorithms.

# Non-Asymptotic Learning Rate

**Proposition 2** Let Assumptions 1-3 hold. Also, let $\rho \in (0,1)$ be a desired probability accuracy, and consider the update rule

$$\mu_{k+1}^i(\theta) = \frac{\prod_{j=1}^n \mu_k^j(\theta)^{[A_k]_{ij}} \ell^i(s_{k+1}^i|\theta)}{\sum_{r=1}^m \prod_{j=1}^n \mu_k^j(\theta_r)^{[A_k]_{ij}} \ell^i(s_{k+1}^i|\theta_r)}$$

Then, the following property is true: for any $\theta \notin \Theta^*$, there is an integer $\boldsymbol{N}(\rho)$ such that, with probability $1-\rho$, for all $k \geq \boldsymbol{N}(\rho)$ there holds

$$\mu_k^i(\theta) \leq \exp\left(-\frac{k}{2}\gamma_2 + \gamma_1\right) \quad \forall i = 1, \ldots, n,$$

$$\boldsymbol{N}(\rho) \triangleq \left\lceil \frac{8\,(\ln\alpha)^2 \ln\left(\frac{1}{\rho}\right)}{\gamma_2^2} \right\rceil,$$

$\alpha$ is a lower bound on the likelihoods $\ell^i$ (see Assum. 3),

$$\gamma_1 \triangleq \max_{\substack{\theta\in\Theta\setminus\Theta^* \\ \theta^*\in\Theta^*}} \left\{ \max_{1\leq i\leq n} \ln \frac{\mu_0^i(\theta)}{\mu_0^i(\theta^*)} \right\} + \frac{12\log n}{1-\lambda} \ln\left(\frac{1}{\alpha}\right)$$

$$\gamma_2 \triangleq \frac{1}{n}\left( \min_{\theta\in\Theta\setminus\Theta^*} \sum_{i=1}^n D_{KL}(f^i \| \ell^i(\cdot|\theta)) - \sum_{i=1}^n D_{KL}\left(f^i \| \ell^i(\cdot|\theta^*)\right) \right)$$

The constant $\lambda$ is related to the mixing properties of the matrices $A_k$ (graphs) and it satisfies the following relations:

□ For general $B$-connected (undirected) graph sequences $\{G_k\}$

$$\lambda = \left(1 - \frac{\eta}{4n^2}\right)^{\frac{1}{B}}.$$

where $\eta$ in the worst case can be of the order $\frac{1}{\mathcal{O}(n)}$.

□ If $B = 1$ and each $A_k$ is the lazy Metropolis matrix, i.e. the stochastic matrix which satisfies

$$[A_k]_{ij} = \frac{1}{2\max\{d_k^i + 1, d_k^j + 1\}} \quad \text{for all } \{i,j\} \in E_k,$$

then

$$\lambda = 1 - \frac{1}{\mathcal{O}(n^2)}.$$

Here, $d_k^i$ is the degree of the node $i$ in graph $G_k$.

# Allowing Conflicting Models

We relax the assumption that

$$\bigcap_{i=1}^{n} \operatorname*{argmin}_{\theta\in\Theta} D_{KL}\left(f_i\left(\cdot\right)\|\ell^i\left(\cdot|\theta\right)\right) \text{ is nonempty} \qquad \text{(non-conflicting models)}$$

Note that

$$\Theta^* = \operatorname*{argmin}_{\theta\in\Theta} \sum_{i=1}^{n} D_{KL}\left(f_i\|\ell^i(\cdot \mid \theta)\right)$$

is non-empty. We will see that the agents' beliefs will vanish for all $\theta \notin \Theta^*$.

We have three different settings:

1. **Time-varying undirected graphs:** $A_k$ is doubly-stochastic with $[A_k]_{ij} > 0$ if $\{i,j\} \in E_k$.
2. **Time-varying directed graphs** $[A_k]_{ij} = \begin{cases} \frac{1}{d_k^j+1} & \text{if } j \in N_k^i \\ 0 & \text{if otherwise} \end{cases}$

   $d_k^i$ is the out degree of node $i$ at time $k$, $N_k^i$ is the set of out-neighbors of node $i$.
3. **Acceleration in static undirected graphs** $\bar{A}_{ij} = \begin{cases} \frac{1}{\max\{d^i,d^j\}+1} & \text{if } \{i,j\} \in E, \\ 0 & \text{if } \{i,j\} \notin E, \end{cases}$

   $d^i$ degree of the node $i$ and $\bar{A} = \frac{1}{2}I + \frac{1}{2}A$,

# Learning Rules

**Time-varying undirected graphs**

$$\mu_{k+1}^{i}(\theta) = \frac{1}{Z_{k+1}^{i}} \prod_{j=1}^{n} \mu_{k}^{j}(\theta)^{[A_k]_{ij}} \ell^{i}\left(s_{k+1}^{i}|\theta\right), \tag{1}$$

$Z_{k+1}^{i}$ is a normalization factor, i.e.,

$$Z_{k+1}^{i} = \sum_{p=1}^{m} \prod_{j=1}^{n} \mu_{k}^{j}(\theta_p)^{[A_k]_{ij}} \ell^{i}\left(s_{k+1}^{i}|\theta_p\right).$$

**Time-varying directed graphs based on recent work with Olshevsky 2015 on subgradient-push algorithms**

$$\mu_{k+1}^{i}(\theta) = \frac{1}{\tilde{Z}_{k+1}^{i}} \left( \prod_{i=1}^{n} \mu_{k}^{j}(\theta)^{[A_k]_{ij} y_k^j} \ell_i\left(s_{k+1}^{i}|\theta\right) \right)^{\frac{1}{y_{k+1}^{i}}} \tag{2}$$

$$y_{k+1}^{i} = \sum_{i=1}^{n} [A_k]_{ij} y_k^j$$

The algorithm uses column stochastic matrices and a ratio consensus algorithm (Kempe et all. 2003)

**Acceleration in static graphs based on a recent paper by Olshevsky 2014**

$$\mu_{k+1}^{i}(\theta) = \frac{1}{\widehat{Z}_{k+1}^{i}} \frac{\prod_{j=1}^{n} \mu_{k}^{j}(\theta)^{(1+\sigma)A_{ij}} \ell^{i}\left(s_{k+1}^{i}|\theta\right)}{\prod_{j=1}^{n}\left(\mu_{k-1}^{j}(\theta)\,\ell^{j}\left(s_{k}^{j}|\theta\right)\right)^{\sigma A_{ij}}} \tag{3}$$

$$\widehat{Z}_{k+1}^{i} = \sum_{p=1}^{m} \frac{\prod_{j=1}^{n} \mu_{k}^{j}(\theta_p)^{(1+\sigma)A_{ij}} \ell^{i}\left(s_{k+1}^{i}|\theta_p\right)}{\prod_{j=1}^{n}\left(\mu_{k-1}^{j}(\theta_p)\,\ell^{j}\left(s_{k}^{j}|\theta_p\right)\right)^{\sigma A_{ij}}} \tag{4}$$

where $\sigma = 1 - 2/(9U+1)$ and $U \geq n$. Obtained by applying Nesterov's accelerated algorithm to the consensus setting.

# Acceleration in Static Graphs: Rate Result

**Theorem 1** *Then, the accelerated update rule with the uniform initial condition* $\mu^i_{-1}(\theta) = \mu^i_0(\theta)$ *has the following property: there is an integer* $\boldsymbol{N}(\rho)$ *such that, with probability* $1-\rho$, *for all* $k \geq \boldsymbol{N}(\rho)$ *and for all* $\theta_v \notin \Theta^*$, *there holds*

$$\mu^i_k(\theta_v) \leq \exp\left(-\frac{k}{2}\gamma_2 + \gamma_1\right) \quad \textit{for all } i = 1, \ldots, n,$$

where

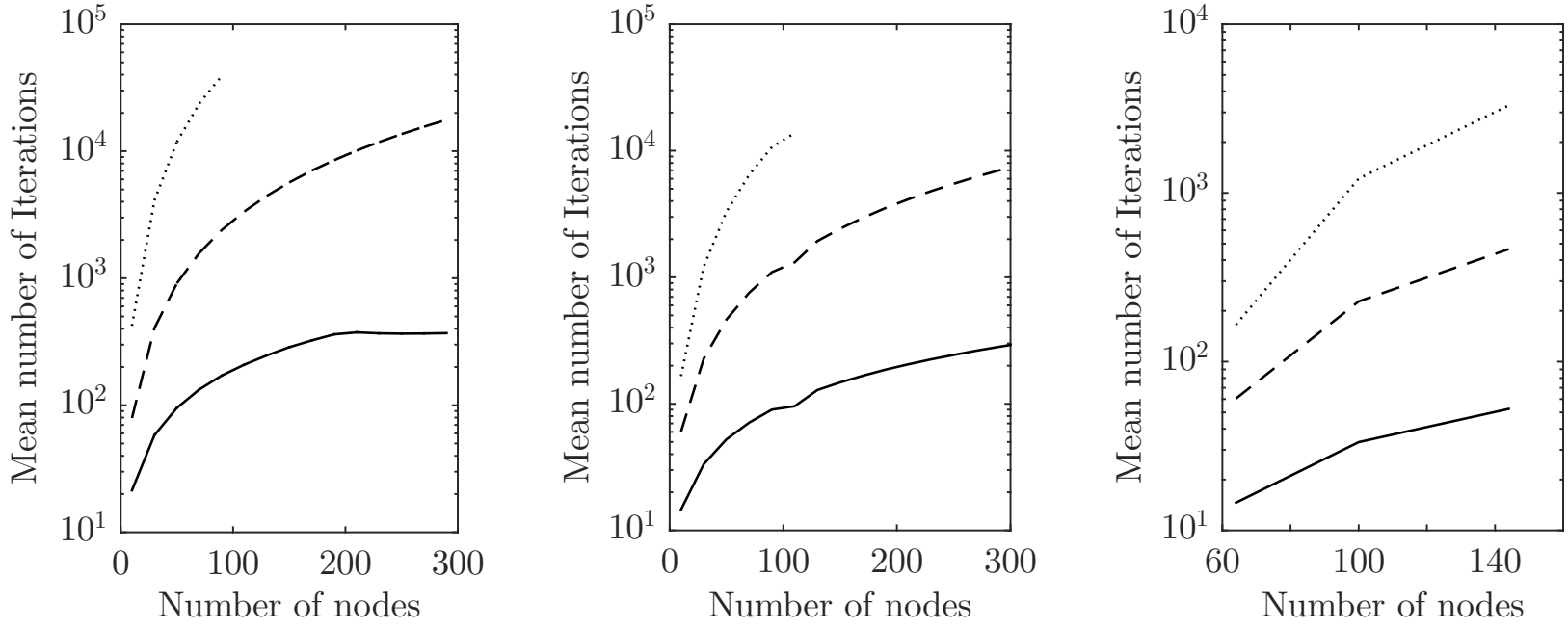
$$\boldsymbol{N}(\rho) \triangleq \left\lceil \frac{48\,(\ln\alpha)^2 \ln\left(\frac{1}{\rho}\right)}{\gamma_2^2} \right\rceil,$$

$$\gamma_1 \triangleq \frac{4\log n}{1-\lambda} \ln\left(\frac{1}{\alpha}\right)$$

$$\gamma_2 \triangleq \frac{1}{n}\left(\min_{\theta \in \Theta \setminus \Theta^*} \sum_{i=1}^n D_{KL}(f^i \| \ell^i(\cdot|\theta)) - \sum_{i=1}^n D_{KL}\left(f^i \| \ell^i(\cdot|\theta^*)\right)\right)$$

with $\alpha$ from the Assumption on likelihoods and $\lambda = 1 - \frac{1}{18U}$ with $U = \frac{1}{O(n)}$.

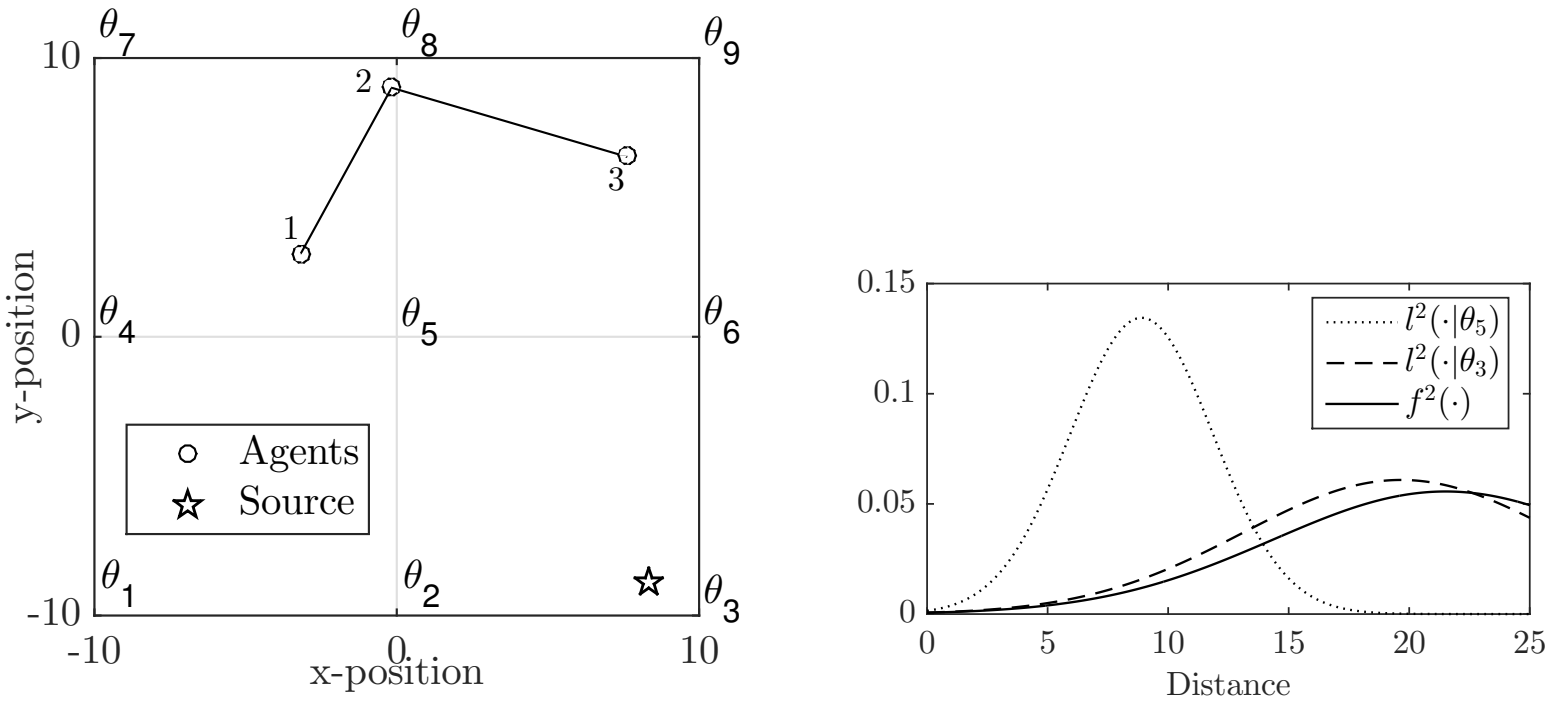
**Note** The index $\boldsymbol{N}(\rho)$ and constant $\gamma_2$ do not depend on the graph structure!




(a) Path Graph (b) Circle Graph (c) Grid Graph

**Figure 1: Empirical mean over 50 Monte Carlo runs of the number of iterations required for $\mu_k^i(\theta) < 0.01$ for all agents on $\theta \notin \Theta^*$. All agents but one have all their hypotheses to be observationally equivalent. Dotted line for the algorithm proposed in [Jadbabaie 2012], Dashed line for the basic algorithm and solid line for the accelerated-procedure.**

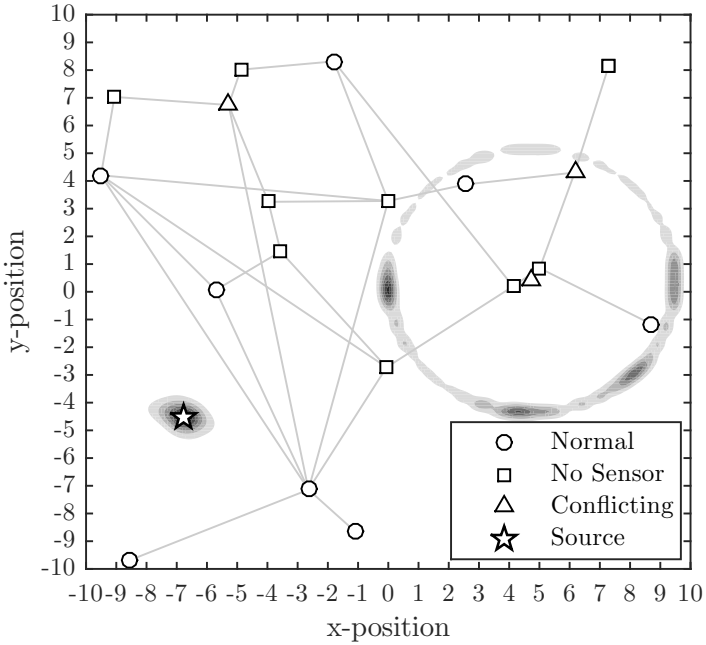
# Distributed Source Localization



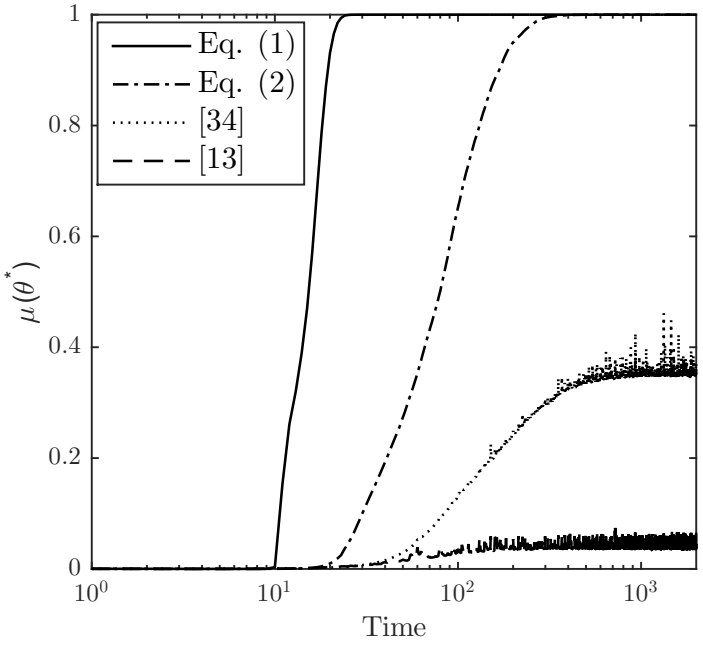

(a) Network of Agents

(b) Hypothesis Distributions

**Figure 2: Figure (a) shows a group of 3 agents in a grid of $3 \times 3$ hypotheses. Each hypothesis corresponds to a possible location of the source. For example, hypothesis $\theta_2$ locates the source at the $(-10, 0)$ point in the plane. Figure (b) shows the likelihood functions for $\theta_2$ and $\theta_5$ and distribution of observations $f^2$ for agent 2.**




(a) Network of Agents




(b) Beliefs on the optimal hypothesis

**Figure 3: Figure (a) shows a network of heterogeneous agents. △ indicates agents whose observations have been modified such that the optimal hypothesis is the $(0, 0)$ point in the grid. □ indicates agents for whom all hypothesis are observationally equivalent (i.e. no data is measured). ○ indicates regular agents with correct observation models and informative hypothesis. Figure (b) shows the belief evolution on the optimal hypothesis $\theta^*$ for different belief update protocols.**

# Related Literature

- Shahrampour, Rakhlin, and Jadbabaie 2014
- Lalitha, Javidi, and Sarwate 2014, 2015
- N. O. Uribe 2014
  **on arxiv**: Fast Convergence Rates for Distributed Non-Bayesian Learning
- N. O. Uribe 2017
  Extended to infinitely many hypotheses - available on **on arxiv**

Thank you